CLAIMS (AS AMENDED)

11. (Restated) A method of identifying from the genomic data of an individual organism a suitable therapy for at least one disease of the organism,

the method particularly serving to identify a relationship between, on a one hand, at least one therapy for at least one disease of an organism, and, on the other hand, genomic data of the organism in the form of two or more alleles and/or SNP pattern(s) of the organism

the method still more particularly serving to determine which of a large number of alleles as variously occur in the genomic data of a large number of individual organisms are, in actual fact, relevant, both individually and in combination, to certain biological and social variables of these organisms, including the efficacy of at least one therapy to at least one disease of these organisms,

the method comprising:

1) constructing a neural network suitable to map (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms as inputs to (ii) historical incidences of responses to therapies for diseases of the individual organisms as outputs; and

2) training the constructed neural network on numerous examples of (i) genomic data as corresponds to (ii) historical incidences of responses to therapies for the diseases of, a multiplicity of individual organisms so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms; and

3) exercising the trained constructed neural network in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between the particular therapy and genomic data, in the form of two or more alleles, of the organisms.

14. (Amended) The method according to claim[s 9, 10,] 11[, 12 or 13]

wherein the training is automated by computerized programmed operations using a genetic algorithm.

15. (Amended) The method according to claim[s 9, 10,] 11[, 12 or 13]

wherein the training is automated by computerized programmed operations using a genetic algorithm reduced in computational complexity by including the steps of:

grouping alleles and/or characteristic SNP patterns into families as are defined by (i) having similar expression patterns or (ii) being turned on and off by another gene, or (iii) both having similar expression patterns and being turned on and off by the same gene; and

starting training of the neural network with the genetic algorithm by using the families so created as single inputs to the neural network, the training with the genetic algorithm continuing repetitively until, families of greater and lessor significance being identified, it becomes computationally possible to train the neural network to genomic data consisting of individual alleles and/or characteristic SNP patterns;

wherein partitioning of all alleles and/or characteristic SNP patterns into families permits training of the neural network in a hierarchy of stages, first to the families and only then to the individual alleles and/or characteristic SNP patterns.

27. (Added) The method according to claim 11 that, at a time before the training of the constructed neural network on numerous examples further comprises:

obtaining, as a first portion of the numerous examples upon which the constructed neural network is trained, (i) genomic data in the form of alleles datums of types taken from a first group consisting essentially of

entire gene families,

specific alleles,

specific base pair sequences,
locations and types of introns, and
nucleotide polymorphism,

plus at least one member of a second, environmental, group consisting essentially of

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and
pharmacokinetic and pharmacodynamic parameters.

28. (Added) The method according to claim 27 wherein the obtaining, as a first portion of the numerous examples upon which the neural network is trained, (i) genomic data in the form of alleles datums from a third, combination genetic and environmental, group consisting essentially of

ethnicity, and
race.

29. (Added) A computerized method of identifying from the genomic data of an individual organism a suitable therapy for at least one disease of the organism, the method comprising:

constructing a neural network relating as inputs (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms to outputs in the form of (ii) historical incidences of responses to therapies for diseases of the same individual organisms; and

training the neural network so constructed on numerous (i) genomic datums, as correspond to (ii) historical incidences of responses to therapies for the diseases, of a multiplicity of individual organisms;

therein making a trained neural network that is fit, and that possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the

organisms; and

exercising the trained constructed neural network in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between a particular therapy and the genomic data, in the form of two or more alleles, of an individual organism;

wherein from the identified relationship it is determinable whether the particular therapy is suitable for the individual organism.

30. (Added) A neural network

suitable to map (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms as inputs to (ii) historical incidences of responses to therapies for diseases of the individual organisms as outputs; and

trained on numerous examples of (i) genomic data as corresponds to (ii) historical incidences of responses to therapies for the diseases of, a multiplicity of individual organisms so as to be fit, and to possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms;

wherein the trained neural network is exercisable in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between the particular therapy and genomic data, in the form of two or more alleles, of the organisms.

31. (Added) The trained neural network according to claim 30

trained by computerized programmed operations using a genetic algorithm.

32. (Added) The trained neural network according to claim 30

trained by computerized programmed operations using a genetic algorithm is reduced in computational complexity by including the

steps of:

grouping alleles and/or characteristic SNP patterns into families as are defined by (i) having similar expression patterns or (ii) being turned on and off by another gene, or (iii) both having similar expression patterns and being turned on and off by the same gene; and

starting training of the neural network with the genetic algorithm by using the families so created as single inputs to the neural network, the training with the genetic algorithm continuing repetitively until, families of greater and lessor significance being identified, it becomes computationally possible to train the neural network to genomic data consisting of individual alleles and/or characteristic SNP patterns;

wherein partitioning of all alleles and/or characteristic SNP patterns into families permits training of the neural network in a hierarchy of stages, first to the families and only then to the individual alleles and/or characteristic SNP patterns.

33. (Added) The trained neural network according to claim 30 that is trained on the numerous examples

obtained, in a first portion, from (i) genomic data in the form of alleles datums of types taken from a first group consisting essentially of

entire gene families,
specific alleles,
specific base pair sequences,
locations and types of introns, and
nucleotide polymorphism,

plus at least one member of a second, environmental, group consisting essentially of

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and

pharmacokinetic and pharmacodynamic parameters.

34. (Added) The trained neural network according to claim 33 that is trained on the numerous examples

further obtained, still in the first portion, from (i) genomic data in the form of alleles datums of types taken from a third, combination genetic and environmental, group consisting essentially of

ethnicity, and

race.

35. (Added) A neural network functioning to identify from the genomic data of an individual organism a suitable therapy for at least one disease of the organism, the neural network

relating as inputs (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms to outputs in the form of (ii) historical incidences of responses to therapies for diseases of the same individual organisms; and

being trained on numerous (i) genomic datums, as correspond to (ii) historical incidences of responses to therapies for the diseases, of a multiplicity of individual organisms; and, by virtue of so relating and of being sot trained

being fit, meaning possessing a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms when exercised in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the training was directed, in order to identify a relationship between a particular therapy and the genomic data, in the form of two or more alleles, of an individual organism;

wherein from exercising of the trained neural network possessing the measure of goodness on the identified relationship it is determinable whether the particular therapy is suitable for the individual organism.

CLAIMS (IN PLAIN TEXT FORM)

11. (Restated) A method of identifying from the genomic data of an individual organism a suitable therapy for at least one disease of the organism,

the method particularly serving to identify a relationship between, on a one hand, at least one therapy for at least one disease of an organism, and, on the other hand, genomic data of the organism in the form of two or more alleles and/or SNP pattern(s) of the organism

the method still more particularly serving to determine which of a large number of alleles as variously occur in the genomic data of a large number of individual organisms are, in actual fact, relevant, both individually and in combination, to certain biological and social variables of these organisms, including the efficacy of at least one therapy to at least one disease of these organisms,

the method comprising:

1) constructing a neural network suitable to map (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms as inputs to (ii) historical incidences of responses to therapies for diseases of the individual organisms as outputs; and

2) training the constructed neural network on numerous examples of (i) genomic data as corresponds to (ii) historical incidences of responses to therapies for the diseases of, a multiplicity of individual organisms so as to make a trained neural network that is fit, and that possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms; and

3) exercising the trained constructed neural network in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between the particular therapy and genomic data, in the form of two or more alleles, of the organisms.

14. (Amended) The method according to claim 11 wherein the training is automated by computerized programmed operations using a genetic algorithm.

15. (Amended) The method according to claim 11 wherein the training is automated by computerized programmed operations using a genetic algorithm reduced in computational complexity by including the steps of:

grouping alleles and/or characteristic SNP patterns into families as are defined by (i) having similar expression patterns or (ii) being turned on and off by another gene, or (iii) both having similar expression patterns and being turned on and off by the same gene; and

starting training of the neural network with the genetic algorithm by using the families so created as single inputs to the neural network, the training with the genetic algorithm continuing repetitively until, families of greater and lessor significance being identified, it becomes computationally possible to train the neural network to genomic data consisting of individual alleles and/or characteristic SNP patterns;

wherein partitioning of all alleles and/or characteristic SNP patterns into families permits training of the neural network in a hierarchy of stages, first to the families and only then to the individual alleles and/or characteristic SNP patterns.

27. (Added) The method according to claim 11 that, at a time before the training of the constructed neural network on numerous examples further comprises:

obtaining, as a first portion of the numerous examples upon which the constructed neural network is trained, (i) genomic data in the form of alleles datums of types taken from a first group consisting essentially of

entire gene families,
specific alleles,
specific base pair sequences,
locations and types of introns, and

nucleotide polymorphism,

plus at least one member of a second, environmental, group consisting essentially of

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and
pharmacokinetic and pharmacodynamic parameters.

28. (Added) The method according to claim 27 wherein the obtaining, as a first portion of the numerous examples upon which the neural network is trained, (i) genomic data in the form of alleles datums from a third, combination genetic and environmental, group consisting essentially of

ethnicity, and
race.

29. (Added) A computerized method of identifying from the genomic data of an individual organism a suitable therapy for at least one disease of the organism, the method comprising:

constructing a neural network relating as inputs (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms to outputs in the form of (ii) historical incidences of responses to therapies for diseases of the same individual organisms; and

training the neural network so constructed on numerous (i) genomic datums, as correspond to (ii) historical incidences of responses to therapies for the diseases, of a multiplicity of individual organisms;

therein making a trained neural network that is fit, and that possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms; and

exercising the trained constructed neural network in respect

of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between a particular therapy and the genomic data, in the form of two or more alleles, of an individual organism;

wherein from the identified relationship it is determinable whether the particular therapy is suitable for the individual organism.

30. (Added) A neural network

suitable to map (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms as inputs to (ii) historical incidences of responses to therapies for diseases of the individual organisms as outputs; and

trained on numerous examples of (i) genomic data as corresponds to (ii) historical incidences of responses to therapies for the diseases of, a multiplicity of individual organisms so as to be fit, and to possesses a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms;

wherein the trained neural network is exercisable in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the neural network was trained, in order to identify a relationship between the particular therapy and genomic data, in the form of two or more alleles, of the organisms.

31. (Added) The trained neural network according to claim 30

trained by computerized programmed operations using a genetic algorithm.

32. (Added) The trained neural network according to claim 30

trained by computerized programmed operations using a genetic algorithm is reduced in computational complexity by including the steps of:

grouping alleles and/or characteristic SNP patterns into

families as are defined by (i) having similar expression patterns or (ii) being turned on and off by another gene, or (iii) both having similar expression patterns and being turned on and off by the same gene; and

starting training of the neural network with the genetic algorithm by using the families so created as single inputs to the neural network, the training with the genetic algorithm continuing repetitively until, families of greater and lessor significance being identified, it becomes computationally possible to train the neural network to genomic data consisting of individual alleles and/or characteristic SNP patterns;

wherein partitioning of all alleles and/or characteristic SNP patterns into families permits training of the neural network in a hierarchy of stages, first to the families and only then to the individual alleles and/or characteristic SNP patterns.

33. (Added) The trained neural network according to claim 30 that is trained on the numerous examples

obtained, in a first portion, from (i) genomic data in the form of alleles datums of types taken from a first group consisting essentially of

entire gene families,
specific alleles,
specific base pair sequences,
locations and types of introns, and
nucleotide polymorphism,

plus at least one member of a second, environmental, group consisting essentially of

diet type,
home region,
occupation,
viral levels,
peptide levels,
blood plasma levels, and
pharmacokinetic and pharmacodynamic parameters.

34. (Added) The trained neural network according to claim 33 that is trained on the numerous examples

further obtained, still in the first portion, from (i) genomic data in the form of alleles datums of types taken from a third, combination genetic and environmental, group consisting essentially of

ethnicity, and

race.

35. (Added) A neural network functioning to identify from the genomic data of an individual organism a suitable therapy for at least one disease of the organism, the neural network

relating as inputs (i) genomic data in the form of two or more alleles and/or SNP patterns of individual organisms to outputs in the form of (ii) historical incidences of responses to therapies for diseases of the same individual organisms; and

being trained on numerous (i) genomic datums, as correspond to (ii) historical incidences of responses to therapies for the diseases, of a multiplicity of individual organisms; and, by virtue of so relating and of being sot trained

being fit, meaning possessing a measure of goodness, to map (i) said genomic data to (ii) incidences of responses to therapies for the diseases of the organisms when exercised in respect of a particular therapy for a particular disease, taken from among the therapies and the diseases to which the training was directed, in order to identify a relationship between a particular therapy and the genomic data, in the form of two or more alleles, of an individual organism;

wherein from exercising of the trained neural network possessing the measure of goodness on the identified relationship it is determinable whether the particular therapy is suitable for the individual organism.

NEURAL-NETWORK-BASED IDENTIFICATION, AND APPLICATION, OF GENOMIC INFORMATION PRACTICALLY RELEVANT TO DIVERSE BIOLOGICAL AND SOCIOLOGICAL PROBLEMS, INCLUDING DRUG DOSAGE ESTIMATION

REFERENCE TO RELATED APPLICATIONS

The present application is a continuation-in-part of U.S. patent application serial number 09/451,249 filed November 29, 1999, for NEURAL NETWORK DRUG DOSAGE ESTIMATION to inventors including the inventors of the invention of the present application. The contents of the related patent application are incorporated herein by reference.

TABLE OF CONTENTS

BACKGROUND OF THE INVENTION

1. Field of the Invention

At an abstract level, the present invention concerns the relationship between (i) genomic data and (ii) disease, and also between (i) genomic data and (iii) disease therapy(ies) -- also known as pharmacogenomics --, as such relationships (i)-(ii) and (i)-(iii) are illuminated by use of neural networks -- neural networks being an extremely powerful mathematical tool preferably exercised in a powerful computer.

In more concrete terms, the present invention generally concerns the (i) identification of genomic data that is relevant in a practical sense to some particular biological or sociological problem afflicting or besetting some type(s) of organism(s), and

the (ii) use of the relevant genomic data so identified so as to select and predict therapy(ies), and any adverse risks and/or consequences thereof, for some particular biological or sociological problem(s) of some particular organism(s).

In still more precise terms, the present invention particularly concerns the selection and training of neural networks for the (i) identification of those particular alleles and/or Single Nucleotide Polymorphism (SNP) patterns within the genomic information of an organism, preferably a human, that are practically relevant to some particular biological or sociological problem afflicting or besetting the organism, most commonly the problem(s) of human disease(s), and, separately, the (ii) use of alleles and/or SNP patterns identified relevant to some disease to predict each of the efficacy, side effects, and expected results of some particular therapy(ies) for some particular patient (who has particular alleles and SNP patterns) in respect of the alleles and/or SNP patterns of this particular patient.

Finally, the present invention concerns a powerful new technique for realizing solutions of neural networks.

2. Description of the Prior Art

The following sections 2.1 through 2.4 are substantially identical to the same sections within the aforementioned related patent application serial number 09/451,249, and discuss prior art relevant to this, as well as the predecessor, invention. They are included within the present specification for sake of completeness. Following sections 2.5 and 2.6 are, however, of unique relevance to the present invention.

2.1 Drug Dosage Estimation by Drug Developers and Physician Practitioners

Many ailments exist in society for which no absolute cure exists. These aliments include, to name a few, certain types of cancers, certain types of immune deficiency diseases and certain types of mental disorders. Although society has not found an

absolute cure for these and many other types of disease, the use of drugs has reduced the negative effects of these disorders.

Generally the developers of drugs have two goals. First, they try to alter the drug user's biochemistry to correct the physiological nature of the illness. Second, they try to reduce the drug's negative side effects on the user. To accomplish these goals, drug developers utilize time consuming and increasingly complex methods. These expensive efforts yield an extremely high cost for many drugs.

Unfortunately, when these costly drugs are distributed they are usually accompanied by only a crude system for assisting a doctor in determining an appropriate drug dosage for a patient. For instance, the annually printed *Physician's Desk Reference* summarizes experimentally determined reasonable drug dosage ranges found in the research literature. These ranges are general. The same dosage range is commonly given for all patients.

Other publications exist which provide general methods to assist a doctor in determining an appropriate dosage. These references and manuals are not, however, directed towards providing a precise dosage range to match a specific patient. Rather, they provide a broad range of dosages based on an averaging of characteristics over an entire population of patients. The correlations between distinguishing patient characteristics and actual required dosages are never obtained, even in the original research.

Faced with the task of minimizing side effects and maximizing drug performance, doctors sometimes refine the dosage they prescribe for a given individual by trial and error. This method suffers from a variety of deleterious consequences. During the period that it takes for trial and error to find an optimal drug dosage for a given patient, the patient may suffer from either (i) unnecessarily high levels of side effects or else (ii) low or totally ineffective levels of relief. Furthermore, the process wastes drugs, because it either prescribes a greater amount of drug than is needed or prescribes such a small amount of drug that it

does not produce the desired effect. The trial and error method also unduly increases the amount of time that the patient and doctor must consult.

2.2 The Need for Drug Dosage Optimization

The past few decades have produced research identifying numerous factors that influence the clinical effects of medication. Age, gender, ethnicity, weight, diagnosis and diet have all been found to influence both the pharmacokinetics and pharmacodynamics of drugs. As a result, it is now acknowledged that women, minorities, and the elderly often require considerably lower doses of some medications than their male Caucasian counterparts. Furthermore, it is possible that patient variables have potentially varying strengths of influence for each case, and each drug.

For example, weight may be of greater importance than age for a Caucasian male while the converse may be true for an African American female. See Lawson, W. B. (1996). The art and science of psychopharmacotherapy of African Americans. Mount Sinai Journal of Medicine, 63, 301-305. See also Lin, K. M., Poland, R. E., Wan, Y., Smith, M. W., Strickland, T. L., & Mendoza, R. (1991). Pharmacokinetic and other related factors affecting psychotropic responses in Asians. Psychopharmacology Bulletin, 27, 427-439. See also Mendoza, R., Smith, M.W., Poland, R., Lin, K., Strickland, T. (!991). Ethnic psychopharmacology: The Hispanic and Native American perspective. Psychopharmacology Bulletin, 27, 449-461. See also Roberts, J., & Tumer, N. (1988). Pharmacodynamic basis for altered drug action in the elderly. Clinical Geriatric Medicine, 4, 127-149. See also Rosenblat, R., & Tang, S. W. (1987). Do Oriental psychiatric patients receive different dosages of psychotropic medication when compared with Occidentals? Canadian Journal of Psychiatry, 32, 270-274. See also Dawkins, K., & Potter, Z. (1991). Gender differences in pharmacokinetics and pharmacodynamics of psychotropics: Focus on women. Psychopharmacology Bulletin, 27, 417-426.

A recent study by Lazarou and colleagues [Lazarou J, Pomeranz

BH, Corey PN. *Incidence of adverse drug reactions in hospitalized patients: a meta-analysis of prospective studies.* JAMA. 1998;279:1200-1205.] noted that in hospitalized patients, the overall incidence of adverse drug reactions (ADRs) was approximately 6.7%. The incidence of fatal ADRs was about 0.32%. In 1994 alone, it is estimated that 2,216,000 hospitalized patients experienced serious ADRs and 106,000 patients had fatal ADRs. ADRs resulting in part from the variability in individual drug response, rank between the 4th and 6th leading causes of death in the United States. Underdosing, overdosing, and misdosing of medications cost the United States more than $100 billion a year.

Pharmacogenomics has the potential to improve drug safety by addressing the issue of why individuals metabolize drugs differently. Informing prescribers of who will metabolize a drug slowly or quickly can optimize drug dosing, improve clinical outcomes, and decrease health costs. [Valdes R. Introduction. Pharmacogenetics in Patient Care Conference. American Association of Clinical Chemistry. Chicago, Ill; Nov 6, 1998.]

Currently, the large number of potentially interacting variables to consider, in addition to the wide therapeutic windows of many drugs (including psychotropic drugs) have resulted in prescribing practices that rely mainly upon trial-and-error and the experience of the prescribing clinician.

The compensation process can be quite lengthy while drug consumers experiment with varying dosages. New methods are needed to reduce the time to compensation for patients (including psychiatric patients), thus alleviating their suffering more quickly as well as reducing the cost of hospitalization. The optimization of drug dosages would also help avoid unnecessarily high dosages, reducing the severity of the many side effects that typically accompany such medications and increasing the likelihood of long-term compliance with the prescribed regimen.

For decades, researchers have recognized the need for finding new methods of accounting for inter-individual differences in drug response. See, for example, Smith, M., & Lin, K. M. (1996); A

biological, environmental, and cultural basis for ethnic differences in treatment; In P. M. Kato, & T. Mann (Eds.), Handbook of Diversity Issues in Health Psychology (pp. 389-406); New York: Plenum Press; and also Lenert, L., Sheiner, L., & Bläschke, T. (1989). Improving drug dosing in hospitalized patients: automated modeling of pharmacokinetics for individualization of drug dosage regimens; Computational Methods in Programs Biomedical, 30, 169-176.

However, a practical solution to tailoring drug regimens has yet to be implemented on a widespread basis.

2.3 Existing Pharmacological Software

Pharmacological software currently in use attempts to provide guidelines for drug dosages, but most software programs merely access databases of information rather than compute drug dosages. At best, these databases rely upon existing research that groups subjects in a few gross categories (e.g., the elderly, or children), and they usually do not include information regarding such relevant characteristics as weight or ethnicity.

The few analytical software products that make use of computer algorithms base their recommendations primarily upon blood plasma concentrations of the drug of interest. See, for example, Tamayo, M., Fernandez de Gatta, M., Garcia, M., & Dominguez, G. (1992); Dosage optimization methods applied to imipramine and desipramine in enuresis treatment; Journal of clinical pharmacy and therapeutics, 17, 55-59; and also Lacarelle B., Pisano P., Gauthier T., Villard P.H., Guder F., Catalin J., & Durand A. (1994); Abbott PKS system: a new version for applied pharmacokinetics including Bayesian estimation; International Journal of Biomedical Computing, 36, 127-30.

Although these methods have met with some success in research, there are several major drawbacks to their implementation. The necessity for constant blood draws for each patient being monitored hinders their practicality in the clinical setting. Furthermore, the limitations of the algorithms used allow modeling of no more

than a few select characteristics at a time, thus ignoring all others. Finally, the models inherently comprise a single algorithm.

However, various drugs have been demonstrated to exhibit quite different response curves. Most new methods use a Bayesian model, which allows for the incorporation of individual response characteristics. See, for example, Tamayo, et al., op. cit. and also Kaufmann G.R., Vozeh S., Wenk M., Haefeli, W.E. (1998). Safety and efficacy of a two-compartment Bayesian feedback program for therapeutic Tobramycin monitoring in the daily clinical use and comparison with a non-Bayesian one-compartment model; Therapeutic Drug Monitoring, 20, 172-80. Even so, the user must first select one rigid modeling equation.

2.4 Present Use of Neural Networks in the Health Sciences

Neural networks will be seen to be used in the present invention. Neural networks have had some, limited, application in the Health Sciences.

Recent research has begun to demonstrate that the flexibility of neural networks in trying a variety of algorithms reduces the margin of error in prediction of blood plasma levels. See Brier, M.E., & Aronoff, G.R. (1996); Application of neural networks to clinical pharmacology; International Journal of Clinical Pharmacology and Therapeutics, 34, 510-514.

The past two to three years have produced a proliferation of studies in the application of neural nets to clinical pharmacology. For example, neural networks are now being used to automate the regulation of anesthesia. See Huang, J.W., Lu, Y.Y., Nayak, A., Roy, R.J. (1999); Depth of anesthesia estimation and control; IEEE Trans Biomedical Engineering, 46, 71-81.

Neural networks are used to determine optimal insulin regimens. See Trajanoski, Z., & Wach, P. (1998); Neural predictive controller for insulin delivery using the subcutaneous route; IEEE Trans Biomedical Engineering, 45, 1122-1134; and also Ambrosiadou, B.V., Gogon, G., Maglaveras, N., Pappas, C. (1996); Decision

support for insulin regime prescription based on a neural net approach; Medical Information, 21, 23-34.

Neural networks are even used to predict clinical response to other medications. See Brier, M.E., et. al., op. cit. and also Bourquin, J., Schmidli, H., van Hoogevest, P., Leuenberger, H. (1997); Application of artificial neural networks (ANN) in the development of solid dosage forms; Pharmacology Development Technology, 2, 111-21.

However, few, if any, prior art references consider the influence of ethnicity. And none known to the inventors envision the comprehensive neural network optimization that will seen to be the subject of the present and related inventions.

The full potential of neural network applications in medicine has yet to be realized, but their growing popularity has resulted in more sophisticated methodology. For example, a genetic algorithm was used to reduce the number of variables required for the training of a neural net in the prediction of patient response to the drug Warfarin. See Narayanan, M.N., & Lucas, S.B. (1993); A genetic algorithm to improve a neural network to predict a patient's response to Warfarin; Methods in Information Medicine, 32, 55-58.

However, most current models used in research are dated and not as efficient as those yet to be publicized -- such as the preferred Levenberg-Marquardt technique used in the present and related inventions, as is explained in detail hereinafter. Furthermore, although genetic algorithms have recently been used in the neurocomputing field to optimize network architectures, these research techniques have yet to be translated to the medical community or to medical applications (as is the subject of the present invention). (NOTE: "Genetic algorithms" as applied to neural networks has nothing to do with genes, and alleles . The phrase "genetic algorithm" is applied in the Darwinian sense, meaning that application of the algorithm serves to identify and make a superior neural network architecture).

2.5 The Motivation for, and Difficulties of, Associating the Genomic Data of an Individual Patient With the Clinical Response(s) to be Expected from the Patient

The present invention will be seen to concern the use of data regarding alleles , both in groups of organisms including men, and for specific organisms or men.

Tabletop screening (with a "bio-chip") of an individual's genome for the identification of a few percent of their alleles is presently (circa 2000) available. The human genome has been announced to have been completely sequenced in this year 2000. In 3-5 years, we expect bio-chips (or families thereof) that can scan an individual's genome for the identification of all of their alleles to become commercially available. The technology will exist to determine an individual unique SNP map. The focus of genomic research will then shift (and is already shifting) to emphasize bioinformatics: how to use the newly discovered clinical genomic data to do useful things.

A major problem with the current state of the field of bioinformatics is that it lacks practical algorithms for extracting from a given genome sufficient relevant information to be of practical use as applied to any of an assortment of biological and sociological problems. The field can only identify individual (or perhaps pairs of) statistically significant alleles that predict a problematic variable value (such as a high risk for breast cancer or Parkinson's disease).

The goals for the end-user are (i) to deliver methods that predict such variables, and, if possible (ii) to predict how therapy, primarily drugs, might beneficially be administered in consideration of the particular alleles of a particular individual. This is a daunting task in which rigor is lacking. It is one thing to say: "This alleles is detected present; based on my experience or inclination as a physician administer this drug." It is another thing to mathematically irreducibly prove that there is some sound factual basis for the prescribed drug therapy. We teach a general procedure for implementing such methods below. Our methods consist

of two parts: 1) identification of relevant alleles combinations and 2) clinical variable prediction given an individual's alleles

Extensive efforts are underway worldwide in diverse locations attempting to associate a person's genetic makeup with, inter alia, the person's susceptibility to disease. These efforts do not, to the best knowledge of the inventors, employ neural networks -- as will seen to be the case with the present invention.

2.6 The Difficulty of Applying a Neural Network to Genomic Data

Neural networks are understood to be powerful problem solving tools for isolating and identifying complex relationships -- exactly the kind of relationships that are believed, and that have been in minute fraction preliminarily identified, between the genomic makeup of an organism and the organism's susceptibility to certain disease(s), probable response(s) to the disease(s), and probable response(s) to any administered therapy(ies) for the disease(s) (if any such exist). Why then have not neural networks been applied to genomic data?

The reason is that the data space (the genome, or even parts thereof) is overwhelmingly large for the tool (the neural network) as implemented on present day (circa 2000) computers (including supercomputers). In order to use a neural network on such an immense data space as the genome is has heretofore been necessary to "guess" which portion of the genome contains the patterns of relevance, and commence neural-network-based analysis on but a minute fraction of the total genome. Since the relationship between genomic coding and disease is presently (circa 2000) very poorly understood for humans, no attempt, let alone any successful attempt, to employ neural networks for identification of the relationship between alleles and/or SNP patterns and disease has not, to the best knowledge of the inventors, yet been reported.

The present invention will be seen to overcome this significant problem by use of two new methods of training a neural network called "householding" and -- as the more important innovation of widespread applicability beyond the genome -- "GA

rolling".

SUMMARY OF THE INVENTION

The present invention contemplates the use of neural networks -- being an extremely powerful mathematical tool preferably exercised in a powerful computer -- in the (i) identification of genomic data that is relevant in a practical sense to some particular biological or sociological problem afflicting or besetting some type of organisms, and, also, the (ii) use of the relevant genomic data so identified so as to select and predict therapy(ies), and any adverse risks and/or consequences thereof, for some particular biological or sociological problem of some particular organism. When, as is most common, the organisms are humans, then the neural-network-based methods of the present invention are most commonly used to (i) identify genomic data in the form of alleles and/or Single Nucleotide Polymorphism (SNP) patterns, that are relevant to human disease(s), and, further, (ii) to predict the efficacy, side effect(s) and response(s) of an individual human patient to a particular therapy(ies) in respect of the genomic data -- the alleles and/or SNP patterns -- of the individual human.

In more precise terms, the present invention firstly contemplates the selection and training of neural networks for (i) the identification of those particular alleles and/or Single Nucleotide Polymorphism (SNP) patterns within the genomic information of an organism, preferably a human, that are practically relevant to some particular biological or sociological problem afflicting or besetting the organism, most commonly the problem of human disease. In accordance with the present invention, this identification is done with and by a neural network -- being an extremely powerful mathematical tool -- that is exercised -- at least in the matter of the human genome -- in a powerful computer accessing a large amount of genomic data in order

to powerfully discern relationships that are presently (circa 2000) substantially unknown, and very difficult to even recognize, let alone to define with mathematical rigor, by any known present techniques.

Also in more precise terms, the present invention secondly, further, contemplates the (ii) practical application of the identified alleles and/or SNP patterns so as to predict the clinical response(s) of some organisms of genomic commonality, and of some particular individual organism -- most commonly men that are alike in respect of the alleles and/or SNP patterns of interest, and of an individual man -- to some stimulus -- particularly drugs -- in consideration of the possession (or lack thereof) of the identified alleles and/or SNP patterns by the genomically common organisms (the like men), or by the particular organism (the individual man). In accordance with the present invention, this prediction also is done with, and by, a neural network.

In realizing these applications the present invention generally teaches (i) the training of neural networks at a first time so as to identify -- out of a vast number of alleles and SNP patterns present in a genomic sequences of each of a large number of individual organisms -- those particular alleles and/or SNP patterns that are relevant in a practical sense to some particular biological or sociological problem afflicting or besetting the organisms, and (ii) the use of neural networks so trained ("trained neural networks") at a second time so as to predict the clinical response of some particular individual organism to some stimulus, particularly drugs, in consideration of the particular organism's possession (or lack thereof) of the identified alleles and/or SNP patterns.

The present invention still further contemplates two new methods of training a neural network. The first method, applicable to genomic data, is called "householding". This method limits the amount of relevant genes by considering (as inputs to the neural network model) only those genes whose expression is similar. In

other words, genes are grouped into families based upon whether they are "on" or "off" at the same time (if this information is known *a priori*). If two or more genes are on or off at the same time, then there is a high probability that they are related, or both are controlled by a third gene. This statistical technique is called "householding", the "householded" genes being treated as a single input to the neural network. This process reduces the amount of data that has to be gathered for use, and the required size of the neural network (which size is related to solution complexity, and time).

The second, and likely more important, method is called "CA rolling". In this method a genetic algorithm (GA) is used to combine ("roll up") a number of inputs to a map into a single input. We use this technique because we suspect that there is approximate symmetry in the genomic inputs, so that their values can be interchanged with little effect on the outputs. This technique dramatically decreases the computational burden placed on the mapping function, which yields improved accuracy. The GA rolling process is more completely explained hereinafter.

1. Identifying the Alleles and Single Nucleotide Polymorphism (SNP) Patterns Relevant in a Practical Sense to Diseases

The present invention contemplates new, neural-network-based, method of identifying those particular alleles and/or SNP patterns -- out of a vast number of alleles and SNP patterns present in the genomic sequences of each of a large number of individual organisms -- that are relevant in a practical sense to some particular biological or sociological problem afflicting or besetting the organisms.

For example, the organisms of primary interest are normally humans. The problem afflicting the humans is most commonly a disease -- by way of example one specific form of cancer, and by way of further example breast cancer. Genomic data as includes, most typically, some hundreds or thousands of alleles and SNP patterns expressed in, most typically, some hundreds or thousands

of genes, is available on a large number of humans as are both afflicted and not afflicted with the disease. Some alleles and/or SNP patterns that affect the occurrence of a specific disease, for example breast cancer, may have been identified, and still other relevant alleles and SNP patterns almost certainly remain unidentified. Furthermore, and even without variables of environment, there are strong indications that some combination of alleles and/or SNP patterns is involved in ultimate susceptibility to the particular disease, to the breast cancer. After all, sometimes only some of several people with nearly identical alleles an/or SNP patterns, for example siblings, will contact the disease. Meanwhile, other persons having widely differing profiles of the alleles and SNP patterns identified as significant will all contact the disease. There is great complexity, and attendantly great confusion, in trying to figure out exactly what correlations and combinations of alleles and/or SNP patterns are, and are not, significant to the occurrence (or non-occurrence) of the disease.

To this complexity is brought a modern mathematical method of tremendous power, executed (for the instance of the human genomic database) on computers of considerable power, most commonly supercomputers. The mathematical method is the (i) selection and (ii) training of neural networks, particularly as are exercised, in accordance with the present invention, by a preferred global optimization algorithm. The computerized method can "sort through" to recognizing relationships that are literally "beyond human ken".

The "solution" of the mathematical method is represented by the (i) selected and (ii) trained neural network. No simple "IF... THEN..." expression can embody the knowledge that comes to reside in such a (i) selected and (ii) trained neural network. It is quite literally impossible to state in words exactly what the (selected, trained) neural network is doing (or, more technically, it may be said that the state equation of the neural network transcends concise expression). Once selected and trained, the neural network may be, and is, exercised with but a tiny fraction of the computational power that built it. The software-based,

selected and trained, neural network commonly runs in personal computer in a physician's office.

The selected and trained neural network will supply answers to questions like: What are the alleles and SNP patterns of importance to contacting breast cancer? What is the probability that person possessed of some subset or superset of these important alleles and will contact breast cancer? If a patient already shows the problem -- e.g., breast cancer -- then what is the prognosis of remission? of reoccurrence? of death? What change in this probability, if any, would result if this person's weight was less? Moreover, a properly selected and trained neural network will likely supply a better answer to these (limited) questions than any human physician on earth.

If the answers to the questions posed the selected and trained neural network in respect of the alleles and/or SNP pattern data of an individual patient are that the patient "has small likelihood of any problem", then that can be the end of the inquiry. However, if the answers to the questions posed are that the "patient has high likelihood of contacting a disease, or a protracted and/or more severe evolution of a disease already detected", then the inquiry must go on.

2. Identifying the Alleles and/or Single Nucleotide Polymorphism (SNP) Patterns Relevant in a Practical Sense to Disease Therapies

The present invention further contemplates a new, neural-network-based, method of identifying those alleles and SNP patterns, as variously possessed in part by some members of a large group of individuals, in combination, which are, in combination, important to predict the clinical response of patients to some particular stimulus or stimuli, particularly drugs administered either in prophylaxis, or in response to, disease. That is, a neural network is selected and trained on a large information data base of, preferably, a population of people that both are and that are not sick, and among certain members of which population disease

is and is not arrested and/or cured, to identify which alleles and/or SNP patterns are, in combination, important in a practical sense to any of (i) disease prevention or (ii) disease arrestment or (iii) disease cure responsive to the stimuli (e.g., to the drugs). As well as predicting drug efficacy relational to alleles and/or SNP patterns, adverse drug reactions can also be predicted.

As with identification in the first instance of those alleles and/or SNP patterns as were associated with a disease, a neural network is both (i) selected and (ii) trained to relate (i) identified pre-selected alleles and SNP patterns (as selectively appear in the genomic sequences of each of large number of historical patients) with (ii) the clinical histories of the response of these patients to some particular disease (e.g., breast cancer) in consideration of therapies applied, most commonly drugs. As before, (i) selecting and (ii) training the neural network to the commonly vast historical clinical data, and to some scores or even hundreds of alleles and/or SNP patterns, is a computationally intensive task normally performed over the period of some hours or days on a supercomputer.

Properly performed -- and causal relationships, howsoever complex and permuted, residing somewhere within the data -- the resulting (i) selected, and (ii) trained, neural network will itself be the "synthesis solution". The neural network will itself be the expression of what can be known from the data.

The later use, and exercise, of the neural network -- discussed in the next section -- is only so as to give "answers" for particular questions (i.e., what should be expected from administration of some particular drug) for particular patients (i.e., as are possessed of a particular pattern of alleles and SNP patterns). Notably, the neural network can exercised so as to validate its own performance (or lack thereof). The clinical data for the many patients, and patient histories, can be fed into the (selected, trained) neural network, one patient at a time. Does the neural network accurately predict what historical data shows to have actually happened? A properly selected and trained neural

network is normally much more accurate in its prognostications (for the useful questions that it may suitably answer) than is any human physician. The physician's judgment ultimately controls, but the "advice" of the neural network "solution" constitutes a useful adjunct to the physician's judgment in the considerably complex area of relating a patient's therapy to his or her genetic profile.

3. Identifying From the Alleles and/or SNP Patterns of a Particular Individual the Therapies Relevant in a Practical Sense to the Disease of Prospective Disease of the Individual

It should be understood that such recognition of (i) the alleles and/or SNP patterns pertinent to various diseases, and (ii) the alleles and/or SNP patterns pertinent to various therapies for various diseases, as is accorded by those methods of the present invention described in immediately preceding sections 1 and 2 is of independent importance, and value. For example, recognition of which alleles and SNP patterns are deterministic as to disease occurrence may accord for such genetic alteration as avoids occurrence of the disease in the first place. For example, recognition of which alleles are important to disease therapy(ies) may accord for such improvement in therapy does effectively safely "cure" the disease, making any further inquiry into the alleles and SNP patterns of a particular patient to be irrelevant.

Normally, however, it is expected that telling an individual patient something of the nature that "(i) 60% of women having the identical profile of (by way of arbitrary, fanciful, example) some five alleles possessed by the patient do die of breast cancer save that (ii) a particular leading therapy is capable of putting 40% of breast cancers overall into remission" will be of scant consolation to the patient, nor value to the patient and her doctor. The patient wants to know what can best be done for her individually, with what associated prognosis.

The present invention further contemplates a new, neural-network-based, method of interpreting in a practical sense the impact of identified alleles and/or SNP patterns, in combination,

possessed by some particular individual so as to predict the clinical response of this particular individual to some particular stimulus or stimuli, particularly drugs. That is, a (selected, and trained) neural network is used to predict a particular individual's response to a particular stimulus, normally a drug, in consideration that the particular individual does, or does not, possess some particular allele, or combination of alleles and/or SNP patterns. As well as predicting drug efficacy, adverse drug reactions can also be predicted.

As with identification of the pertinent alleles and SNP patterns in the first instance, a neural network is both (i) selected and (ii) trained to relate (i) identified pre-selected alleles and SNP patterns (as selectively appear in the genomic sequences of each of large number of historical patients) with (ii) the clinical histories of the response of these patients to some particular disease (e.g., breast cancer) in consideration of therapies applied, most commonly drugs. As before, (i) selecting and (ii) training the neural network to the commonly vast historical clinical data, and to some scores or even hundreds of alleles and/or SNP patterns, is a computationally intensive task normally performed over the period of some hours or days on a supercomputer. Properly performed -- and causal relationships, howsoever complex and permuted, residing somewhere within the data -- the resulting (i) selected, and (ii) trained, neural network will itself be the "synthesis solution". The neural network will itself be the expression of what can be known from the data.

The later use, and exercise, of the neural network is only so as to give "answers" for particular questions (i.e., what should be expected from administration of some particular drug) for particular patients (i.e., as are possessed of a particular pattern of alleles or SNPs). Notably, the neural network can exercised so as to validate its own performance (or lack thereof). The clinical data for the many patients, and patient histories, can be fed into the (selected, trained) neural network, one patient at a time. Does the neural network accurately predict what historical data

shows to have actually happened? A properly selected and trained neural network is normally much more accurate in its prognostications (for the useful questions that it may suitably answer) than is any human physician. The physician's judgment ultimately controls, but the "advice" of the neural network "solution" constitutes a useful adjunct to the physician's judgment in the considerably complex area of relating a patient's therapy to his or her genetic profile.

4. Training a Neural Network on the Immense Genomic Data

The present invention contemplates a novel computerized method for processing in a neural network (i) a large amount of genomic data including a large number of genes with (ii) a large number of clinical results in order to train the neural network with a training algorithm to map the genomic data into the clinical results. The method is improved over previous methods of training a neural network in that, before the training begins, the amount of relevant genes are limited by statistical processes so as to consider substantially only those genes with a similar expression is similar. To "limit by statistical properties" simply means that genes are grouped into families based upon *a priori* information as to whether the genes are "on" or "off" at the same time. If two or more genes are on or off at the same time then these two or more genes are treated as a single unit. Alternatively, if these two or move genes are not "on" or "off" at the same time then they are treated separately. This improvement wherein limiting of the number of inputs is realized by grouping of the inputs is called "householding".

This improvement is preferably used as part of training a neural network with a genetic algorithm, or GA, and is more preferably used in the training of a neural network with a genetic algorithm of the rolling type, or a "rolling GA".

This "rolling GA" algorithm is itself novel. In accordance with the present invention, it is a method of adapting a very great number of datums to a much smaller number of inputs to a neural

network during training of the neural network to map its inputs to a small number of outputs. The method requires the availability of a common scalar cost function to measure error on the outputs of a neural network. The method process by processing in the neural network a large number of binary fuzzy inputs to map to neural network outputs, the error of which outputs is measured. In consideration of the measured output errors, a given mapping is "broken up" into (i) a preprocessor that categorizes the inputs and (ii) a secondary mapping with fewer inputs.

Second and subsequent mappings transpire -- each in a neural network for so many times as are required -- until, by hierarchial reduction through intermediate mappings in a tree-structured hierarchy of neural networks, the very great number of datums distributed as inputs among a plurality of leaf node neural networks are mapped in a hierarchy of neural networks until only the very small number of outputs is produced by a final, root node, neural network.

In this hierarchy of mappings all of the very great number of datums having no significance to the final outputs tend to become grouped together as but a single input to the root node neural network, which input is accorded zero weight. In this hierarchy of mappings all of the great number of datums that are, as binary fuzzy inputs, relative to said final outputs tend to be mapped through successive hierarchical stages, or "rolled", from inputs to outputs, and do thus contribute to said final outputs.

The neural network is preferably modeled with a set of architectural mapping parameters that can be optimized by a genetic algorithm.

The method is commonly performed on inputs divided into an arbitrary number of categories, each category containing a finite artificial genome representing the full set of N inputs to the original mapping. The number of inputs N is preferably in the range from 10 to 50, and the number x in the range from 5 to 15.

To recapitulate, the preferred neural network mapping is on (i) inputs that have underdone "householding", meaning that

multiple genes are treated as a single unit, by (ii) use of a Genetic Algorithm (GA) that is "rolled", meaning that mapping transpires in neural networks organized hierarchically in stages so as to relate a typically vast amount genomic data as neural networks inputs to but very little clinical data as the outputs of a final, root node, neural network.

5. The "Rolling Genetic Algorithm" of the Present Invention

In greater detail, and with mathematic rigor, the "rolling genetic algorithm", or "rolling GA", of the present invention may be considered, as applied to genomic data, to be embodied in a method of training a neural network having a multiplicity M of inputs so as to extract information from genomic data having a great multiplicity of N variables, N >> M. Unknown ones and unknown numbers of a majority of which N variables are both irrelevant and non-contributory to information that is extractable as desired output from a trained neural net. The method is thus directed to training a neural network having only M inputs to extract information from N variables, N >> M, where, although many of the N variables are irrelevant or of much lesser relevance than others of the N variables, it is not known which, nor what number, of the N variables are so substantially irrelevant to extracting the information. The method is of the general nature of an exercise of dual strategies of (i) divide and conquer while (ii) suppressing incorporation of substantially irrelevant variables until, finally, a neural network, nonetheless to having only M inputs, is trained to extract information from genomic data having a great multiplicity of N variables where M << N.

In the method a great multiplicity of N genomic variables are organized into M categories, called artificial genes, where M << N;

A same set of N input values are input into each of these M categories as a functional block.

By use of the M artificial genes and the N input values (i) a vector of N values, or weights, is created for each of the M artificial genes, the weights being initially set randomly.

A dot (scalar) product of (i) the N-valued vector with (ii) an input vector of N genomic variables is defined so as to create (iii) one single output value.

A dot product between successive (ii) input vectors each of a successive N genomic variables and (i) the vector of N values that are initially random, is repetitively derived for each of the M functional blocks.

This repetitive derivation -- some M times -- creates a filter vector, or artificial chromosome, of M values, which M values correspond to M genes in the artificial chromosome.

A neural network is used to map the created filter vector, or artificial chromosome, as an input vector so as to calculate a cost output value. This cost output value is a function of how similar the neural network output value is to a desired result. The mapping also takes into consideration how many of the weights in the artificial genes are sufficiently below some predetermined threshold so as to be considered negligible.

A cost output value is optimized so as to create, by modifying the weights of each artificial gene, a particular artificial chromosome which, when fed as an input vector into the mapping of the neural network, causes the output values of the neural network to assume an optimal cost function.

By these steps the number of inputs to the mapping neural net is decreased to M out of the N genomic variables, $M << N$. Thus, proceeding from the great multiplicity of N genomic variables, (i) those variables which have greatest relevance to the optimal output of the mapping neural net are preferentially selected while (ii) those variables which have least relevance to the optimal output of the mapping neural network are preferentially discarded. Furthermore, the great multiplicity of N genomic variables are divided into M categories, or artificial chromosomes, having similar functionality.

The optimizing of the vector inputs to the M functional blocks which have assigned to them a unique output value preferably transpires by use of a genetic algorithm.

The method is in particular useful to identify a statistically significant group of N genomic datums in the form of alleles and/or SNP patterns as these genomic datums affect given clinical results, which group is generally known as a clinically relevant alleles combination and/or characteristic SNP pattern as the case may be, proceeding from genomic data of N variables.

6. Objectives of the Present Invention

Accordingly, one objective of the present invention is the identification of those alleles and SNP patterns that are associated, in a practical sense, with each of an immense number of biological and social variables. In so doing the present invention will employ powerful automated techniques based on (i) programmed neural networks (ii) selected and trained in powerful computers.

Another objective of the present invention is to predict at least one clinical variable of an individual patient in respect of alleles and/or SNP pattern data of the individual patient. To do so, the present invention will teach the training of a neural network, and the clinical use of the neural network so trained.

Still another objective of the present invention is to screen an individual patient for expected reaction to a drug in respect of the alleles and/or SNP pattern data of the individual patient. To do so, the present invention will again teach the training of a neural network, and the clinical use of the neural network so trained.

Yet still another objective of the present invention is to predict an optimal drug dosage for an individual patient in respect of alleles and/or SNP pattern data of the individual patient. To do so, the present invention will yet again teach the training of a neural network, and the clinical use of the neural network so trained.

These and other aspects and attributes of the present invention will become increasingly clear upon reference to the following drawings and accompanying specification.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1a is a diagram of the motivation for identification of functional alleles families, such as transpires in the present invention.

Figure 1b is a flowchart of the preferred method of identifying clinically relevant alleles combinations in accordance with the present invention.

Figure 1c is a flowchart of the structure of neural network training routine in accordance with the present invention.

Figure 1d is a block diagram of a typical mapping neural net in accordance with the present invention.

Figure 1e is a flow chart of a typical genetic algorithm in accordance with the present invention.

Figure 2 is a flow chart of the method of predicting clinical variables given genomic data in accordance with the present invention.

Figure 3 is a diagram of the preferred genomic methods of screening patients for clinical drug use in accordance with the present invention.

Figure 4a is a diagram of the preferred "GA rolling" sub-process of the present invention.

Figure 4b is a diagram of the application of the preferred "GA rolling" sub-process of the present invention applied to an infeasible initial mapping problem.

Figure 4c is a diagram illustrating an individual category and its genes.

Figure 4c is a diagram illustrating the mapping used by the preferred genetic algorithm of the present invention.

Figure 4d is a diagram illustrating the preferred method of using the preferred genetic algorithm of the present invention.

Figure 5a is a diagram illustrating preliminary constructs in the use of functional genomic categorizations for predicting drug interactions in accordance with the present invention.

Figure 5b is a flow chart illustrating intermediate